

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-020248
 (43)Date of publication of application : 21.01.2000

(51)Int.Cl. G06F 3/06
 G06F 12/16

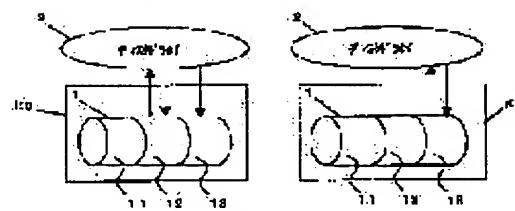
(21)Application number : 10-196766 (71)Applicant : NIPPON DIGITAL KENKYUSHO:KK
 (22)Date of filing : 26.06.1998 (72)Inventor : YOKOYAMA MASAHIKO

(54) DISK FAULT RECOVERING METHOD, DISK DRIVER, AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a disk fault recovering method, a disk driver and a disk which are applicable to even an inexpensive stand-alone type computer device by dividing the disk into areas and automatically duplexing data.

SOLUTION: The disk driver 2 sends a command to a disk device and receives a stylus from the disk device to control access to the disk 1. The disk 1 is divided (partitioned) into a system block where an OS and application programs are stored, a data block 12 where data are stored, and a backup block 13 where the same data as the data written to the data block 12 are stored and the disk driver 2 actualizes mirroring for even a disk device which does not have disk array structure by writing the data to the data block 12 and backup block 13 to make a fault recovery of data.



LEGAL STATUS

[Date of request for examination] 22.06.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-20248

(P2000-20248A)

(43) 公開日 平成12年1月21日 (2000.1.21)

(51) Int.Cl. ⁷	識別記号	F I	テマコード [*] (参考)
G 06 F 3/06	3 0 4	G 06 F 3/06	3 0 4 B 5 B 0 1 8
	5 4 0		5 4 0 5 B 0 6 5
12/16	3 1 0	12/16	3 1 0 J

審査請求 未請求 請求項の数5 FD (全11頁)

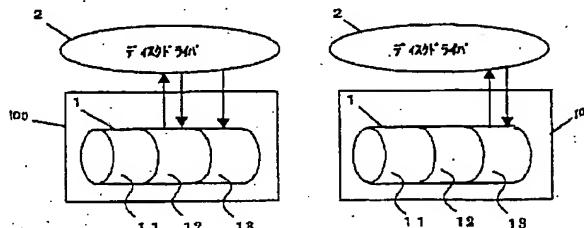
(21) 出願番号	特願平10-196766	(71) 出願人	591008605 株式会社日本デジタル研究所 東京都江東区新砂1丁目2番3号
(22) 出願日	平成10年6月26日 (1998.6.26)	(72) 発明者	横山 昌彦 北海道江別市西野幌120-5 株式会社日本デジタル研究所札幌研究開発センター内
		(74) 代理人	100072383 弁理士 永田 武三郎
			Fターム(参考) 5B018 GA02 GA04 GA06 HA03 KA01 KA02 KA14 KA18 MA12 QA01 QA16 RA14 5B065 BA01 CC03 EA24 EA31 EA35

(54) 【発明の名称】 ディスクの障害回復方法、ディスクドライバおよび記録媒体

(57) 【要約】

【課題】 ディスクを領域分割して自動的にデータの二重化を行うことにより低価格なスタンダードアローンタイプのコンピュータ装置にも適用可能なディスクの障害回復方法、ディスクドライバおよびディスクの提供。

【解決手段】 ディスクドライバ2はディスク装置にコマンドを与え、ディスク装置からステータスを受取ってディスク1へのアクセス制御を行なう。ディスク1はOSやアプリケーションプログラムを格納するシステムブロック1 1、データを格納するデータブロック1 2およびデータブロック1 2に書込まれたと同一のデータを格納するバックアップブロック1 3に領域分割(パーティション)されており、ディスクドライバ2はデータをデータブロック1 2とバックアップブロック1 3に書込むことによってディスクアレイ構造を有しないディスク装置についてもミラーリングを実現し、データの障害回復を行なう。



1

2

【特許請求の範囲】

【請求項1】 ディスク領域を同容量の2ブロックを含む3ブロックに分割し、前記同容量の2ブロックに同一データを書き込み、前記同容量のブロックのうち一方にディスクエラーが発生したときにはエラー未発生のブロックへのアクセスを行なうことを特徴とするディスクの障害回復方法。

【請求項2】 コンピュータ装置に内蔵または外付け接続して動作するディスク装置へのアクセス制御を行なうプログラムからなるディスクドライバであって、ディスクへのデータの書き込みリクエストがあったとき、同容量の2ブロックを含む3ブロックに分割されたディスクの前記同容量の一方のブロックへのデータの書き込み終了後、更に、同一データを前記同容量の他のブロックに書き込むための書き込みコマンドをディスク装置に発行してディスク装置に書き込みを行なわせ、

ディスクへのデータの読み出しリクエストがあったとき、前記同容量のブロックのうち所定のブロックから専らデータの読み出しを行ない、前記所定のブロックにエラーが発生したときには前記同容量の他のブロックからデータの読み出しを行なうことを特徴とするディスクドライバ。

【請求項3】 前記同容量の2ブロックにデータの書き込みを行なう場合に、前回アクセスを終了したセクタを保持し、今回の書き込みを前記セクタを含むブロックからの書き込み制御を行なうことを特徴とする請求項2記載のディスクドライバ。

【請求項4】 前記同容量の他のブロックをバックアップ用ブロックとし、バックアップ用ブロックにはアプリケーションによるリクエストを禁止することを特徴とする請求項2または3記載のディスクドライバ。

【請求項5】 コンピュータ装置に内蔵または外付け接続して動作するディスク装置によってデータの書き込み／読み出しが行なわれる記録媒体であって、

前記ディスクの領域を同容量の2ブロックを含む3ブロックに分割し、前記同容量の2ブロック中の一方をデータの読み出しを他の一方のブロックより優先するデータブロックとし、前記同容量の2ブロック中の他方を前記データブロックと同一データを書き込むバックアップブロックとし、前記3ブロック中の同容量でないブロックをプログラムまたはインターフリタデータの読み出し／書き込みを行なうシステムブロックとして分割したことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はディスクの障害回復方法、ディスクの障害回復方法を実現するディスクドライバ（ディスク制御プログラム）およびディスク（記録媒体）に関する。

【0002】

【従来の技術】 従来、速度改善のアイデアとして複数台のディスクをアレイ状に構成し、シーク時間や回転待ち時間を短縮することによって高速化を図るディスクアレイ方式があったが、ディスクアレイ方式では複数のディスクをアレイ状に構成する上で必然的に部品点数が増大し、この結果信頼性の低下という問題が生じた。このような問題点を解決すべく冗長情報をアレイの構成に含めることによって、ディスク装置へのアクセスの高速化と障害防止対策を兼ね備えた技術としてRAID (Redundant Arrays of Independent (Inexpensive) Disks) がある。

(1) RAID 1

RAIDは、データを分割して複数のディスクに格納し、高速アクセスと高信頼性を実現する技術であり、分割格納と障害対策の方式の違いによって幾つかのレベルに分類されており、米国カルフォルニア大学バークレー校のパターソン博士（Dr.Patterson）らの論文（Patterson,D.A,Gibson,G,Katz,R.H, "A Case for Redundant Arrays of Inexpensive Disks(RAID)," Report No.UCB/CSD 87/391, Computer Science Division, University of California Berkeley, Dec 1987）では5種類のRAIDが定義されている。なお、その中ではRAID 0は含まれていないが、その有用性からRAID 0はRAIDの一種として考えられているので、現在ではRAIDは次に示すようにRAID 0～RAID 5の6レベルに分類されている。そして、レベルによってディスク容量の使用効率やデータアクセスの効率などが異なっている。

- ① RAID 0 ストライピング
- ② RAID 1 ミラーリング（二重化）
- ③ RAID 2 ストライピング（バイト）とハミングコードによるECC
- ④ RAID 3 ストライピング（バイト）とパリティドライブ固定
- ⑤ RAID 4 ストライピング（ブロック）とパリティドライブ固定
- ⑥ RAID 5 ストライピング（ブロック）とパリティドライブ分散

例えば、RAID 1ではディスクの二重化（ミラーリング）を行なう。すなわち、図9（a）、（b）に示すように全く同じ内容のデータを2台（2台以上でもよい）のディスク94-1, 94-2に書き込む。これにより、あるディスクが故障しても正常な方のディスクを用いて処理を実行できる。なお、RAID 1には図9（a）に示すようにホストコンピュータ91のディスクドライバ92などのソフトウェアによる処理で2重化を行なうソフトウェアRAIDと、図9（b）に示すようにディスク装置のコントローラ93などのハードウェアを利用して2重化を行なうハードウェアRAIDがある。

【0003】 どちらも書き込みは図10（a）に示すように両方のディスク44-1, 44-2に同時に行ない、

読み出しは図10(b)に示すように一方のディスク44-1から行なう。図10(c)に示すように一方のディスク44-2が故障した場合は、もう一方のディスク44-1のみで動作する。すなわち、図9(a)のソフトウェアRAIDではディスクドライバ42が同一データをディスク44-1, 44-2に同時に書込んで2重化を行なう。また、図9(b)のハードウェアRAIDではホストコンピュータ41からのリクエストやコマンドを受取ったコントローラ43が同一データをディスク44-1, 44-2に書込んで同時に2重化を行なう。

【0004】[シングルディスクシステム] シングルディスクシステムは、図11に示すように1台のディスクを用いてデータの書込み/読み出しを行なうものであり、ディスクに書込まれたデータの信頼性向上は別段考慮されていない。シングルディスクシステムではホストコンピュータ91側のデバイスドライバ92'がディスク装置にコマンドを(書込み/読み出し命令)発行し、データの転送を行なってディスク94へのデータの書込み/読み出しを行なう。

(2) その他

その他のディスクアレイの障害回復方法として特開平6-230903号公報には、障害発生時に空き領域を障害回復領域として確保することによりディスクアレイの障害回復を行なう技術が開示されている。上記公報記載の技術はRAID3~5(ストライピングとパリティドライブを用いる方式)における障害回復の改善技術であり、RAID3, 4, 5のいずれかを適用したディスクアレイに障害が発生した場合に、正常ドライブの空き領域を障害回復領域として確保し、障害ドライブのデータを回復領域に分散格納すると共に回復されたデータの2次パリティを生成して、新たなドライブの障害時におけるデータ消失を防止しようとするものである。

【0005】

【発明が解決しようとする課題】近年、Windows95やWindowsNT(いずれも、米国マイクロソフト社の登録商標)等のOS(オペレーティングシステム)の採用により、ハードディスク(HDD)に対するアクセス量、データ量が増大し、それに伴いハードディスクの障害頻度が増大し、それによるデータ損失等の被害が増大している。ディスクの障害には書込み/読み出しヘッド等の機械的機構部分の故障によるディスククラッシュ、ディスクに傷などがついたりした場合のセクターエラー、ノイズによる誤書き込みによるデータエラー等がある。

【0006】RAID1では、ディスクにハードウェア上の故障が生じた場合に重要なデータを保護できるが、2台(または2n台)のディスクに同じデータを書込むので、シングルシステムに比べコストがかかるという問題点、使用可能なディスクの容量が実質的に2分の1以下になってしまいうといふ問題点、上述したソフトウェア

RAID(図9(a))ではデータ書込み時に2回ディスクにアクセスしなければならないので、通常アクセスの2倍以上書込み時間がかかるという問題点、ハードウェアRAID(図9(b))ではコントローラのコストがさらに必要となるという問題点があった。

【0007】また、シングルディスクシステムでは、ディスクに冗長性がないので、ディスクにエラーが生じた場合にデータの復元ができないという欠点があった。

【0008】また、特開平6-230903号公報記載のディスクアレイの障害回復方法では、障害が発生した場合に一時的空き領域を障害回復領域として確保するものであり、障害発生時のペナルティが大きいという問題点がある。

【0009】本発明は、上記問題点および欠点を解決するためになされたものであり、ディスクアレイでの高速アクセスと高信頼性を実現するためにディスクアレイの各ディスクにデータを分割して格納するRAIDとは異なり、ディスクアレイ構成を有しないディスク装置について、ディスクの故障発生時に高い確率でデータを回復できる、ディスクの障害回復方法、ディスクドライバおよびディスクの提供を目的とし、ディスクを領域分割して自動的にデータの二重化を行なうことにより低価格なスタンダードアローンタイプのコンピュータ装置にも適用可能なディスクの障害回復方法、ディスクドライバおよびディスクの提供を目的とする。

【0010】

【課題を解決するための手段】上記課題を解決するため、第1の発明のディスクの障害回復方法は、ディスク領域を同容量の2ブロックを含む3ブロックに分割し、前記同容量の2ブロックに同一データを書込み、前記同容量のブロックのうち一方にディスクエラーが発生したときにはエラー未発生のブロックへのアクセスを行なうことを特徴とする。

【0011】また、第2の発明のディスクドライバは、コンピュータ装置に内蔵または外付け接続して動作するディスク装置へのアクセス制御を行なうプログラムからなるディスクドライバであって、ディスクへのデータの書込みリクエストがあつたとき、同容量の2ブロックを含む3ブロックに分割されたディスクの同容量の一方のブロックへのデータの書込み終了後、更に、同一データを同容量の他のブロックに書込むための書込みコマンドをディスク装置に発行してディスク装置に書込みを行なわせ、ディスクへのデータの読み出しリクエストがあつたとき、同容量のブロックのうち所定のブロックから専らデータの読み出しを行ない、所定のブロックにエラーが発生したときには同容量の他のブロックからデータの読み出しを行なうことを特徴とする。

【0012】また、第3の発明は上記第2の発明のディスクドライバにおいて、同容量の2ブロックにデータの書込みを行なう場合に、アクセスを終了したセクタを保

持し、今回の書込みを前記セクタを含むブロックからの書込み制御を行なうことを特徴とする。

【0013】また、第4の発明は上記第2または第3の発明のディスクドライバにおいて、前記同容量の他のブロックをバックアップ用ブロックとし、バックアップ用ブロックにはアプリケーションによるリクエストを禁止することを特徴とする。

【0014】また、第5の発明の記録媒体は、コンピュータ装置に内蔵または外付け接続して動作するディスク装置によってデータの書込み／読み出しが行なわれる記録媒体であって、ディスクの領域を同容量の2ブロックを含む3ブロックに分割し、同容量の2ブロック中的一方をデータの読み出しを他の一方のブロックより優先するデータブロックとし、同容量の2ブロック中の他方をデータブロックと同一データを書込むバックアップブロックとし、3ブロック中の同容量でないブロックをプログラムまたはインターフリタデータの読み出し／書込みを行なうシステムブロックとして分割したことを特徴とする。

【0015】

【発明の実施の形態】<ディスクパーテーション>図1は、本発明のディスクの障害回復方法を適用可能なディスクの領域構成の一実施例を示す図であり、本発明ではディスクを3つのブロック（パーテーション）にわけて構成する。

【0016】図1で、ディスク1は制御プログラムやアプリケーションプログラムおよびユティリティプログラム等のプログラムやインターフリタコードを格納したり、プログラムの実行時に中間値等を格納したりするシステムブロック11と、データファイルが格納されるデータブロック12、データブロック12に書込んだデータと同じデータを格納するバックアップブロック13を有している。なお、ディスク1は、本実施例ではハードディスク（HDD）を用いているがこれに限定されない。

【0017】これらブロック11～13はディスクの物理的連続領域（通常はセクタ単位）をメーカー側で予め分割して出荷するか、或いはディスクのフォーマット時に割り付けられる（データブロック12とバックアップブロック13は同じ大きさが割当てられる）。また、ディスク1上にディスク1のエラー発生箇所およびその種類（書込みエラー、読み取りエラー等の別）を書込むログ領域を確保するようにしてもよい（本実施例では、ログ領域をコンピュータ装置本体（ホスト）に設けたフラッシュメモリ等の不揮発性メモリに確保している）。

【0018】なお、本実施例では上述のようにコンピュータ装置が1台のディスク装置100を内蔵または外付けしている場合を例とし、1つのディスクをシステムブロック、データブロック、およびバックアップブロックの3ブロックに分割したが、コンピュータ装置が2台ま

たはそれ以上の台数のディスク装置を内蔵または外付けできる場合に、2台またはそれ以上のディスク装置で駆動される各ディスク（記録媒体）の領域を連続的領域とみなして3ブロックに分割してもよい。また、コンピュータ装置が内蔵または外付けするディスク装置の台数nを、 $n = a + 2m$ とするとき、a台をシステムブロックとし、m台をデータブロックとし、他のm台をバックアップブロックとみなすことができる（a, m ≥ 1）。

【0019】<ディスクバックアップシステムの概要>図2は、本発明のディスクの障害回復方法の説明図であり、（a）はアプリケーションプログラムの通常動作中のディスクへのアクセス状態を示し、（b）はデータブロックに障害が生じたときの動作状態を示す図である。以下の説明では、1台のディスク装置がコンピュータ装置（以下、ホストと記す）に内蔵または外付けされた場合を例とするが、上述したように複数のディスク装置がホストコンピュータに内蔵または外付けされていてもよい。すなわち、ディスクアレイではデータは分割されて各ディスクに離散的に格納されるが（並列的格納）、本発明ではデータは連続したブロック内に書込まれ、読み出される（直列的格納）。

【0020】図2で、ディスクドライバ2はディスク装置100にコマンドを与え、また、ディスク装置100やOSとステータスを授受しながらディスク1へのアクセス制御を行なうディスク制御用プログラム群で構成されており、ディスクバックアップ用に本発明に基づく次のような機能を有するディスクバックアッププログラム群を含んでいる。なおディスクドライバをディスクバックアップモジュールを含むハードウエア（ディスクコントローラ）で構成することもできる）。また、ディスクドライバ2はホスト側にあってもよいし、ディスク装置側にあってもよい。

【0021】また、図2で、ディスク1へのアクセスは書込み時にはデータブロック12およびバックアップブロック13の双方に書込み、正常動作では読み出しへデータブロック12から行なう。一方のブロックにエラー（セクタエラー等）が生じた場合には、他の一方のブロックで読み出し／書込みを行なう。

【0022】[ディスクドライバの機能概要]

1. アクセス頻度の高いブロック、すなわち、システムブロック11にはミラーリング（データの二重化）を行なわぬ、システムブロック12に書込まれるデータ以外のデータ（アプリケーションプログラムにより作成されり、更新されるデータファイルに書込まれたり／読み出されたりするデータ）のミラーリングを行なう。
2. 上記データのミラーリングのために、ディスク1上にデータブロック12と同サイズ（容量）のバックアップブロック13を確保する。
3. ユーザ（すなわち、ユーザ用アプリケーションプログラム或いはユーザの操作）によってはバックアップブ

ロック13へのアクセスは読み出し、書き込み共にできない。

4. ミラーリングはファイル単位ではなくデータ単位で行なう。

5. OSからのデータ読み出しのためのアクセス要求、すなわちデータリードアクセスリクエストがあった場合には、データブロック12のデータの読み出しを最優先とする。

6. OSからのデータ書き込みのためのアクセス要求、すなわちデータライトリクエストがあった場合には、高速化を図るために前回アクセスしたセクタに近いブロックからアクセスする。例えば、バックアップブロック13にあるセクタにアクセスしていた場合には今回のアクセスはバックアップブロック13の当該セクタに近い位置(続くセクタ)にデータを書込んでから、データブロック12に同一データを書きみに行く。

7. データブロック12またはバックアップブロック13のいずれか一方にエラーが発生した場合には、シングルシステム(図11)の場合のようにOSにエラー発生ステータス(エラー発生状態を示すフラグ)を返さず、ディスクドライブ2は書き込みの場合には現在書き込もうとしているデータをエラー未発生ブロックに書き込むようにしてから、読み出しの場合にはデータブロック12から読み出そうとしたデータをバックアップブロック13から読み込むようにしてから、エラー通知(例えば、警告ランプ表示、警告メッセージ表示、警告音出力、警告メッセージ出力等によるユーザへのエラー発生通知)を行なう。

8. データブロック12またはバックアップブロック13のいずれか一方にエラーが発生した場合には、上述したように、書き込みの場合には現在書き込もうとしているデータをエラー未発生ブロックに書き込むようにしてから、以後、書き込みデータの転送をエラー未発生ブロックのみに行なう。また、読み出しの場合にはデータブロック12から読み出そうとしたデータをバックアップブロック13から読み込むようにしてから、以後、読み込みアクセスをバックアップブロック13のみに行なう。

9. エラー発生時のディスクエラーログ(エラー発生日時、エラー発生ディスク、エラー発生セクタ、エラーの種類等を示すディスクエラー情報)は、ホストの不揮発性メモリに格納し、電源が投入時に前回の処理でディスクにエラーが生じている場合に表示する。

10. システムブロック11にエラーが生じた場合には、OSにエラー発生ステータス(エラー発生状態を示すフラグ)を返す。以後の処理はOS側の指示に基づく。

11. データブロック12およびバックアップブロック13の両者にエラーが発生した場合には、OSにエラーステータス(エラー発生状態を示すフラグ)を返す。以後の処理はOS側の指示に基づく。この場合、ディスク

ドライバ2は現在のブロックへのアクセスを禁止しないようにするため(すなわち、現在のブロックへのアクセスをOS側の指示により行なえるようにするため)、現在ブロックをアクセスブロックとしてOSにエラーステータスを返す。

【0023】<ディスクドライバの動作>

[正常動作時] 図3は、ディスク装置100が正常動作(ディスクエラーが発生していないときの動作をいう)を行なっているときのディスクドライバ2の基本的動作の説明図であり、(a)はOSからシステムブロック11へのリード/ライトアクセスまたはデータブロック12からのデータ読み出しのためのデータリードリクエストがあつた場合のディスクドライバ2の基本的動作の説明図、(b)はデータブロック12へのデータライトリクエストがあつた場合のディスクドライバ2のデータアクセス時の基本的動作の説明図である。また、図4はディスクドライバ2のデータアクセス時の基本的動作を示すフローチャートである。以下、図3および図4を基にディスクドライバ2のアプリケーションプログラムによるデータアクセス時の基本的動作について説明する。なお、図3(a)、(b)で矢印の記号S0～S18は図4のフローチャートのステップを意味する。

【0024】図3および図4で、コンピュータ装置でアプリケーションプログラム3からのディスクアクセス要求があると、OS4はディスクドライバ2にディスクアクセスリクエストを出す(S0)。

【0025】ディスクドライバ2はOS4からのディスクアクセスリクエストを受取ると、リクエストを調べ(S1)、それがリードリクエストの場合には読み込み命令およびデータブロック12の読み込み開始セクタおよび転送セクタ数等からなるリードコマンドをディスク装置100に発行し、OS4からのディスクアクセスリクエストがライトリクエストの場合には書き込み命令およびデータブロック12の書き込み開始セクタおよび転送セクタ数等からなるライトコマンドをディスク装置100に発行する(S2)。

【0026】ディスクドライバ2はディスク装置100へのコマンド発行後、割込み待ち処理を行ない、割込み待ちステータス(状態フラグ等)をOS4に渡す(OS4が割込み待ちステータスを受取った場合、OS4はアプリケーションプログラム3にディスク装置100が割込み待ちであることを知らせる。この場合、アプリケーションプログラム3の処理がディスク1への(またはディスク1からの)データ転送を伴わない処理であれば処理の実行は可能となり、そうでない場合にはOS4から転送開始通知を受取るまで、実行待ちとなる)。ディスク装置100は転送割込み可能(アクセス可能)となると、割込みステータスをディスクドライバ2に返す(または、ディスクドライバ2は、ディスク装置100に設けられているステータスバイトを所定タイミングでサー

チし、ステータスバイトの割込み相当ビットがオンになつたか否かを判定する) (S 3)。

【0027】ディスクドライバ2は、上記ステップS 0でOS 4から受取ったディスクアクセシリクエストを調べ(S 4)、ディスクアクセシリクエストがリードリクエストの場合にはOS 4に割込みステータスを返し(S 5)、ディスク装置100にデータ転送許可を出して、ディスク1のデータブロック12からアプリケーション側の読み込み用バッファ(メモリ)へのデータ(読み出しデータ)の転送を行なわせ(S 6)、最終割込み処理終了(=転送終了:転送セクタ数をカウントによる判定)後(S 7)、リクエスト終了処理を行ない(S 8)、OS 4に終了通知を出す(S 9)。

【0028】また、上記ステップS 0でOS 4から受取ったディスクアクセシリクエストが、ライトリクエストの場合にはディスクドライバ2はOS 4に割込みステータスを返し(S 10)、ディスク装置100にデータ転送通知を出して、アプリケーション側の読み込み用バッファ(メモリ)からディスク1のデータブロック12へのデータ(書き込みデータ)の転送を行なわせ(S 1 1)、データブロック12への書き込みが終ると(S 1 2)、最終割込み処理を行なうことなくディスク装置100にバックアップブロック13へのライトコマンド(書き込み命令、バックアップブロック13の書き込み開始セクタ数、転送セクター数等を含む)を発行し(S 1 3)、ディスク装置100がデータの書き込み(転送)可能を意味する割込みステータスが返されると、それをOS 4に渡し(S 1 4)、ディスク装置100にデータ転送通知を出して、アプリケーション側の読み込み用バッファ(メモリ)からデータブロック12に書き込んだ同一のデータ(書き込みデータ)をバックアップブロック13に転送を行なわせる(S 1 5)。

【0029】ディスク装置100へのデータ転送(=バックアップブロック13への書き込み)の最終処理が終了すると(S 1 6)、ディスクドライバ2はリクエスト終了処理を行ない(S 1 7)、OS 4に終了ステータスを渡す(S 1 8)。

【0030】上記動作により、ディスク装置100は正常動作時にはディスク1のデータブロック12からデータを読み出し、書き込みはデータブロック12とバックアップブロック13に同一データを書込む。すなわち、ディスク装置1台を内蔵または外付けする最小ディスク構成のコンピュータ装置でもミラーリングを行なうことができる。

【0031】なお、上記図4の説明ではアプリケーションプログラムによるデータアクセス時の基本的動作について説明したが、OSによるプログラムアクセス時の動作、すなわち、システムブロック11のアクセスについては、上記フローチャートのうち、ステップS 0、ステップS 1 2およびS 1 6を省略すると共に、ステップS 50

1～S 1 1、ステップS 1 6～S 1 8でアクセスするブロックをデータブロック12ではなくシステムブロック13とし、アプリケーションプログラム3をOS 4とし、データをプログラム等とすればよい。

【0032】[エラーが生じた場合の動作] 図5は、ディスク装置100にエラーが発生した場合のディスクドライバ2の基本的動作の説明図であり、(a)はシステムブロック11にエラーが発生した場合のディスクドライバ2の基本的動作の説明図、(b)はデータブロック12またはバックアップブロック13にエラーが生じた場合のディスクドライバ2の基本的動作の説明図である。

(:システムブロックのエラー) 図5または図6で、ディスクドライバ2はディスク装置100から送られるステータスを調べ(T 1)、システムブロック11のエラーが発生した場合にはエラースティタスをOS 4に渡して、エラー回復処理をOS 4に委ねる(T 2)。

【0033】(データブロック、バックアップブロックのエラー) 上記ステップT 1で、データブロック12またはバックアップブロック13にエラーが発生した場合には、データブロック12とバックアップブロック13に共にエラーが発生しているか否かを調べ、データブロック12とバックアップブロック13に共にエラーが発生している場合(データブロック12とバックアップブロック13に同時にエラーが生じた場合と、一方のブロックにエラーが生じた後で他の一方に新たにエラーが生じた場合とを含む)にはT 4に遷移し、データブロック12とバックアップブロック13のいずれか一方にエラーが発生した場合にはT 6に遷移する(T 3)。

【0034】(:データブロックとバックアップブロック共にエラー) 上記ステップT 3で、データブロック12とバックアップブロック13に共にエラーが発生した場合には、エラースティタスをOS 4に渡し、OS 4またはアプリケーションプログラム3にエラー処理を委ねる。この場合、ディスクドライバ2は現在のブロックへのアクセスを禁止しないようにするため(すなわち、現在のブロックへのアクセスをOS側の指示により行なえるようにするため)、現在のブロックをアクセスブロックとしてOSにエラー発生ステータスを返し(T 4)、ユーザへのエラー通知を行なう。実施例ではエラー通知はエラー通知プログラム5が行なっているので、ディスクドライバ2はこの場合エラー通知プログラムに遷移する(T 5)。

【0035】(データブロックまたはバックアップブロックの一方のエラー) 上記ステップT 3で、データブロック12またはバックアップブロック13のいずれか一方にエラーが発生した場合で、リードアクセスの途中でエラーが生じたときにはデータをバックアップブロックから読み出す(データ転送)。すなわち、データブロック12にエラーが生じた場合にはバックアップブロック

13の読み出し開始アドレス(セクタ)に現在の転送セクタ数を加えて読み出し再開アドレスとしてOS4に与え、バックアップブロック13からデータを読み出すようにしてから(T6)、ディスクドライバ2はエラー通知プログラムに遷移してユーザへのリードエラー通知を行なわせる(T7)。また、上記ステップT3で、ライトアクセスの途中にエラーが生じた場合で、エラーがデータブロック12に生じたときにはバックアップブロック同一データを書込むようにバックアップブロック13の書込み開始アドレス(セクタ)をOS4に与えると共に、ライトコマンドをディスク装置100に与えてバックアップブロック13にデータを書き直しを開始させ(T8)、T11に遷移する(T9)。また、ライトアクセスの途中にエラーが生じた場合で、エラーがバックアップブロック13に生じた場合には、書込み中断ステイタスをOS4に渡してライトアクセスを終了させる(T10)。ディスクドライバ2はエラー通知プログラムに遷移してユーザへのライトエラー通知を行なわせる(T11)。

[0036] [ディスクアクセス効率向上動作] 本発明のディスクの障害回復方法では、ディスクドライバ2がデータブロック12とバックアップブロック13に同一データを書込むミラーリングを行なうので、一方のブロックへの書込みを終って、他方のブロックにアクセスを開始する場合に、RAID1のように書込みヘッドをそのつど移動していくはシーケン時間がかかり、1台のディスクを用いる場合にはシングルディスクシステム(図11)の2倍時間がかかることになるが、本発明ではコンピュータ装置の起動時やアプリケーションの実行時にシステムブロック11からのOSやプログラムの読み出しや、設定値の書込み等とデータブロック12からのデータ読み出し時には、後述するようにシングルシステムの場合と同様のアクセスを行ない、ミラーリングによる時間効率の低下を最小限に止めるようにしている。また、ディスク容量についても、システムブロック11を二重化しないことにより、データブロックとバックアップブロックのディスク容量比を高くするようにしている。

[0037] (:データ読み出し動作) 図7は、ディスクドライバ2によるディスク装置100のデータ読み出し動作制御の説明図であり、(a)はディスク装置100が正常動作している場合のシステムブロック11に対する読み出し動作の説明図、(b)はディスク装置100が正常動作している場合のデータブロック12に対する読み出し動作の説明図、(c)はデータブロック12にエラーが生じた場合のバックアップブロック13に対する読み出し動作(エラー発生時)の説明図である。

[0038] 図7(a), (b)に示すようにディスク装置100が正常動作している場合の読み出し動作は、OSからリクエストされたブロックのデータを転送する動作のみであり、この場合のアクセス速度はシングルデ

ィスクシステムの場合と同様となる。また、図7(c)に示すようにデータブロック12にエラーが発生した場合には、データブロック12からの転送はおこなわず、バックアップブロック13から転送するので、この場合のアクセス速度もシングルディスクシステムの場合と同様となる。

[0039] (:データ書き込み動作) 図8は、ディスクドライバ2によるディスク装置100の書き込み動作制御の説明図であり、(a)はディスク装置100が正常に動作している場合で、データブロック12にデータを書き込んだ場合の次の書き込みにおけるアクセスブロックの説明図、(b)はディスク装置100が正常動作している場合で、バックアップブロック13にデータを書き込んだ場合の次の書き込みにおけるアクセスブロックの説明図である。

[0040] 図8(a), (b)に示すようにディスク装置100が正常動作している場合の書き込み動作は、シーク時間(ヘッドの移動時間)の短縮を図るために、シングルシステムの場合と同様に今回のアクセスで最初に書き込むセクタを前回アクセスしたセクタに近いブロックからアクセスすることによって結果としてデータ書き込み時間を短縮する。

[0041] 具体的には、図8(a)でデータ書き込みをデータブロック12から開始すると、データのミラーリングを行なうためにデータブロック12に書き込みを終了した後はバックアップブロック13から同一データの書き込みを開始してバックアップブロック13のあるセクタ(p)で書き込みを終了し、書き込み終了セクタ(p)の位置をメモリに保持する。ここで、次のアクセスでデータブロック12から書き込みを開始するようにするとシーク時間がかかるので、これをほぼゼロにするために前回のアクセスの終了位置であるバックアップブロック13の書き込み終了セクタ(p)に最も近い位置である(p+1)セクタまたは終了セクタ(p)を含むブロック(この場合はバックアップブロック13)から書き込みを開始して、バックアップブロック13への書き込みを終了するとデータブロック12に同一データの書き込みを開始してデータブロック12のあるセクタ(q)で書き込みを終了する。また、次のアクセス動作はデータブロック12の書き込み終了セクタ(q)に最も近い位置である(q+1)セクタまたは終了セクタを含むブロック(この場合はデータブロック12)から書き込みを開始してデータブロック12への書き込みを終了するとバックアップブロック13への同一データの書き込みを開始する。このようなアクセス動作を繰り返すことにより、シーク時間が短縮される。

[0042] すなわち、バックアップブロック13からデータブロック12にヘッドを移動させる時間をtとすると、アクセス毎にデータブロック12からの書き込みを開始する方式では、n回アクセスを行なうと(n-1)

t バックアップブロック 13 からデータブロックのシーク時間がかかる。これに対し、本発明の書き込み動作のように前回のアクセス終了位置に近いセクタから今回のアクセスを開始する方式では、前回のアクセス終了位置に近いセクタから今回データを書き込む位置までのヘッダの移動時間を t' とし、 n 回アクセスを行なうと $(n-1) t'$ のシーク時間がかかるが、 $t' << t$ であるから、 $(n-1) t' << (n-1) t$ となり、アクセス開始シーク時間が大幅に節約できる。

【0043】また、エラー発生時のエラー情報を格納するログ領域を不揮発性メモリに設けることにより、次回電源立ち上げ時にディスクにアクセスしなくともエラーの発生をユーザに通知できる。

【0044】

【発明の効果】上記に述べたとおり、本発明によれば、ハードディスクに生じる障害からデータを保護する際、ハードディスクに障害が生じることを回避或いは防止することを直接の目的とするのではなく、ハードディスクに障害が生じることを自明の事象として受け入れ回復することを目的としてデータのミラーリングを行なっているのでディスクエラーの発生時に高い確率でデータを回復できる。特に、低価格なスタンダードアローンタイプのコンピュータ装置、例えば、1台のディスク装置を接続するか内蔵して稼働可能に構成されたコンピュータ装置にも適用可能である。

【0045】ディスクアレイではデータは分割されて各ディスクに離散的に格納されるが、本発明ではデータは連続したブロック内に書き込まれ、読み出されるので、ディスクドライバの構成が RAID の場合より簡素化できる。また、複数のディスクがホストコンピュータに内蔵または外付けされていてもよい。

【0046】前回のアクセス位置を保持して今回の書き込み動作を行なうようにしているので、ミラーリング時の高速化が実現できる。

【図面の簡単な説明】

【図1】本発明のディスクの障害回復方法を適用可能なディスクの領域構成の一実施例を示す図である。

【図2】本発明のディスクの障害回復方法の説明図である。

【図3】ディスク装置が正常動作を行なっているときのディスクドライブの基本的動作の説明図である。

【図4】ディスクドライバのデータアクセス時の基本的動作を示すフローチャートである。

【図5】ディスク装置にエラーが発生した場合のディスクドライバの基本的動作を示す説明図である。

【図6】ディスクドライバのデータアクセス時の基本的動作を示すフローチャートである。

【図7】ディスクドライバによるディスク装置のデータ読み出し動作制御の説明図である。

【図8】ディスクドライバによるディスク装置の書き込み動作制御の説明図である。

【図9】RAID1によるディスクの二重化(ミラーリング)方式の説明図である。

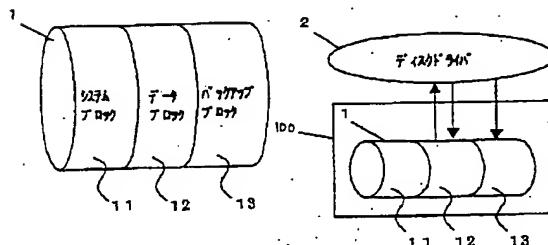
【図10】RAID1によるディスク障害時のデータ回復動作の説明図である。

【図11】シングルディスクシステムの説明図である。

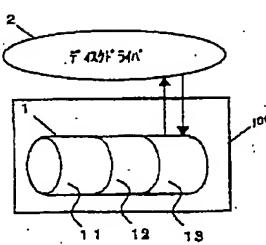
【符号の説明】

- 1 ディスク(記録媒体)
- 2 ディスクドライバ
- 3 アプリケーションプログラム(アプリケーション)
- 11 システムブロック
- 12 データブロック
- 13 バックアップブロック
- 100 ディスク装置

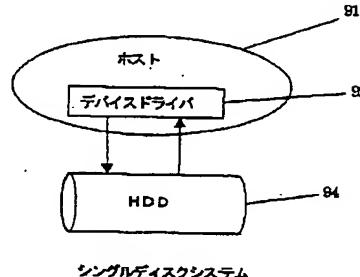
【図1】



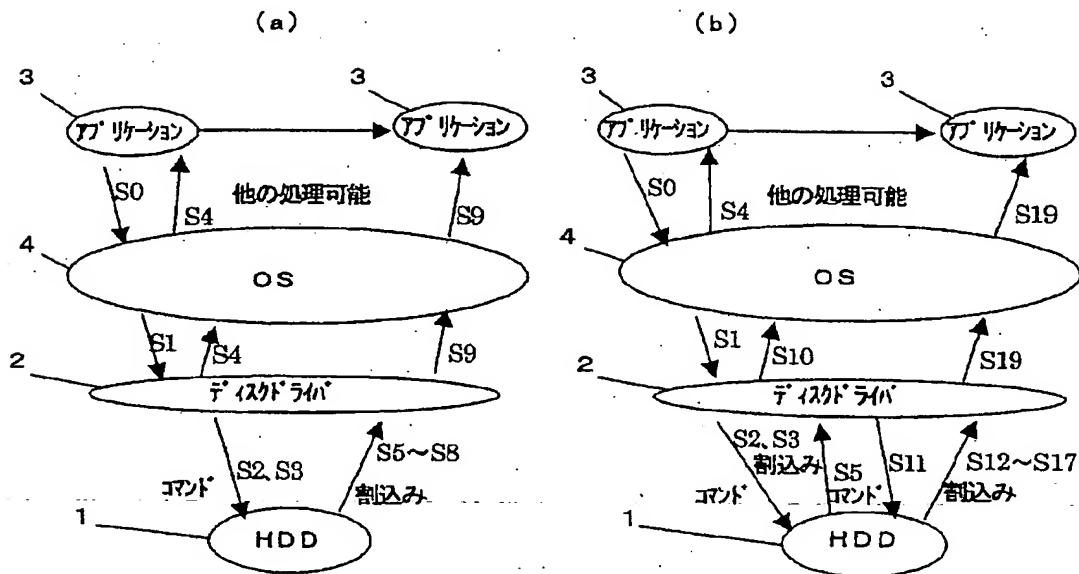
【図2】



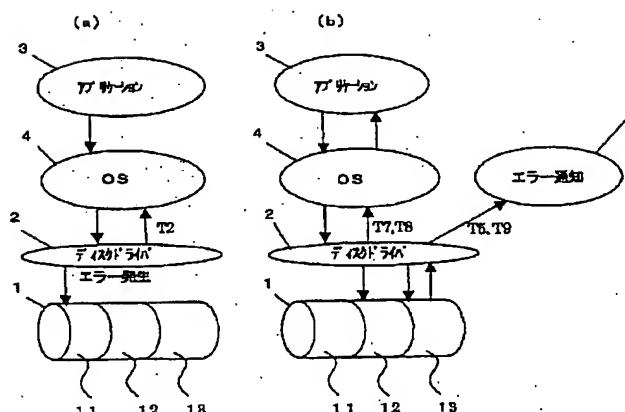
【図11】



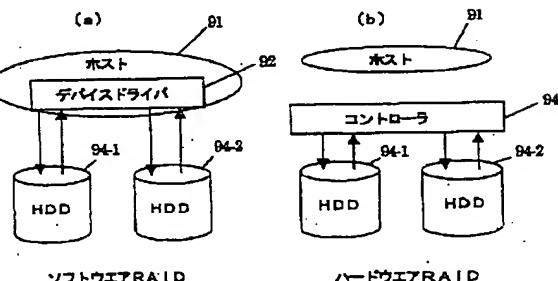
【図3】



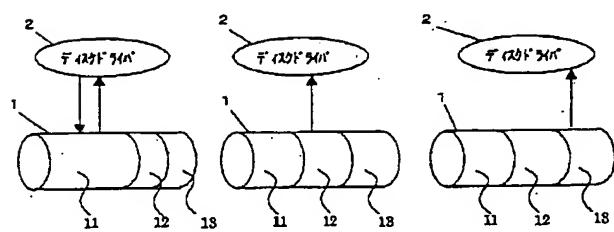
【図5】



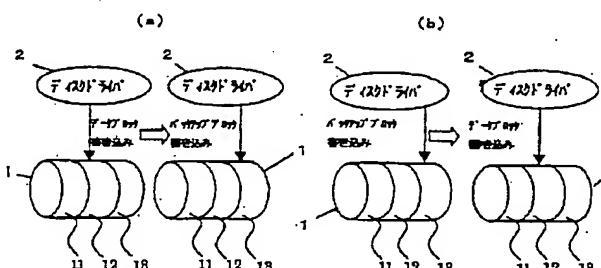
【図9】



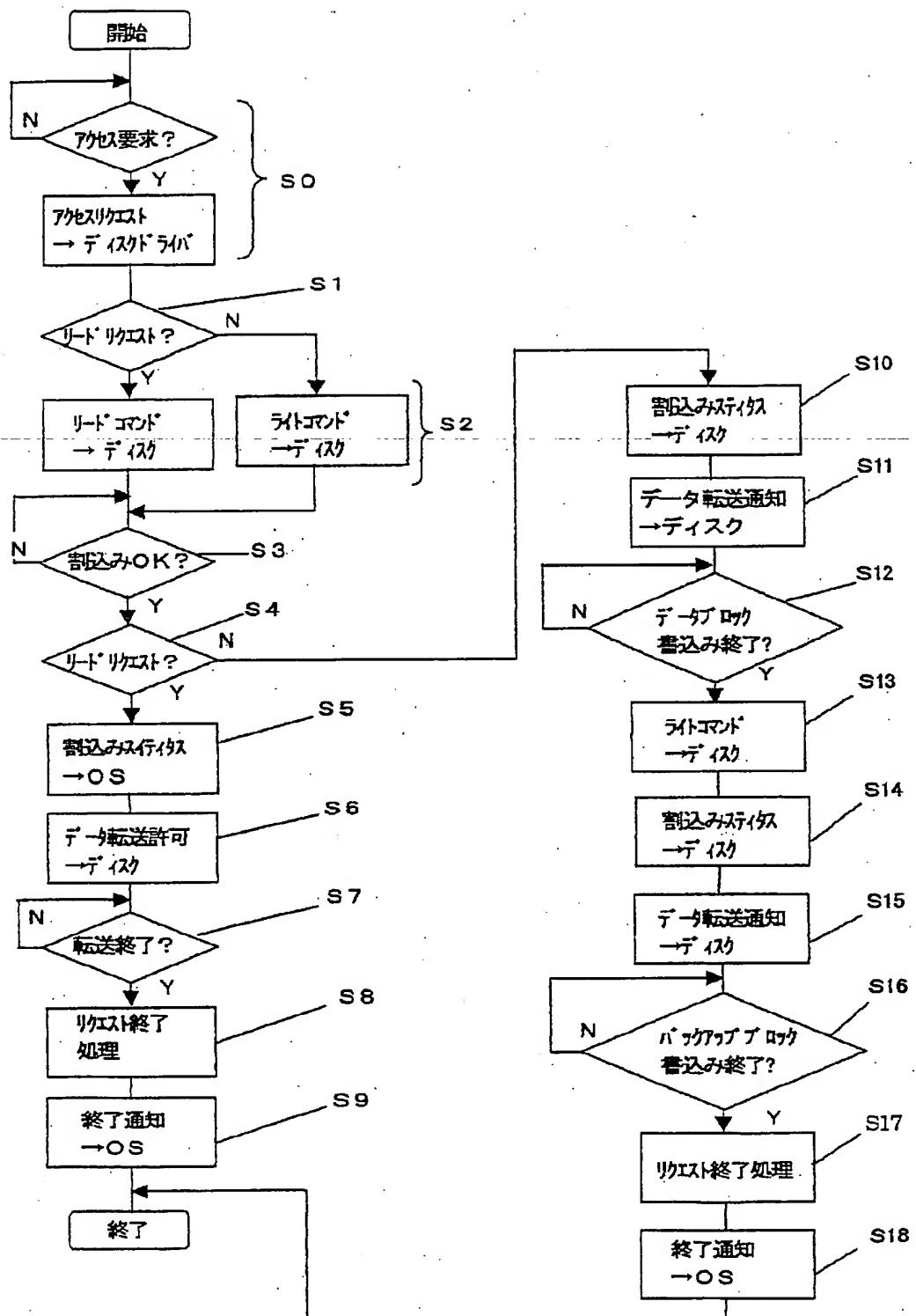
【図7】



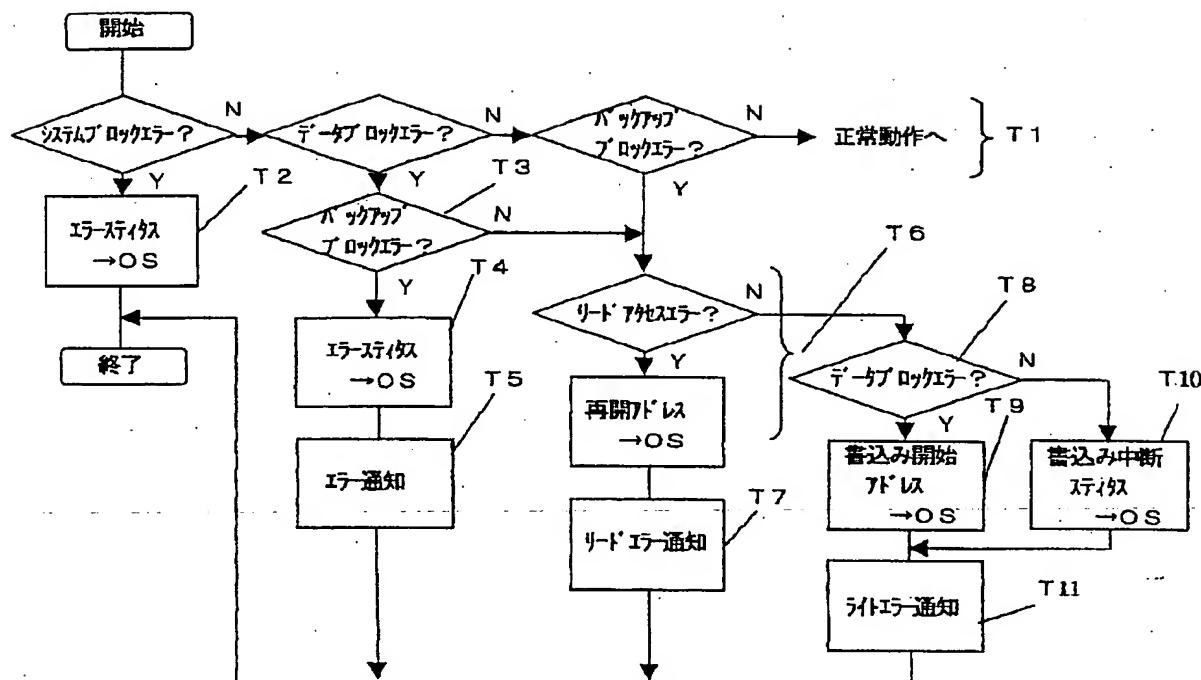
【図8】



【図4】



【図6】



【図10】

